GENERALIZED QUASI-LIKELIHOOD RATIO STATISTICS FOR MULTIVARIATE TIME-VARYING COEFFICIENT REGRESSION MODELS

by

Yi Liu

A dissertation submitted to the faculty of
the University of North Carolina at Charlotte
in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in
Applied Mathematics

Charlotte

2013

Approved by:

____
Dr. Jiancheng Jiang

____
Dr. Yanqing Sun

____
Dr. Weihua Zhou

____
Dr. Aixi Zhou

ABSTRACT

YI LIU. Generalized Quasi-Likelihood Ratio Statistics for Multivariate Time-varying Coefficient Regression Models.
(Under the direction of DR. JIANCHENG JIANG)

Generalized likelihood ratio statistics have been a generally applicable method for testing nonparametric hypotheses about nonparametric functions.It has been widely used in many research areas. It was proposed in Fan,Zhang and Zhang (2001) and been extended to additive models in Fan and Jiang (2005). In this dissertation, I extend their work to multivariate case, my aim is to construct some test statistic to test whether the coefficients are indeed constants or some specific parametric/nonparametric functions for the time-varying coefficient model and the asymptotic null distribution of the proposed test statistic is independent of the nuisance parameters.

ACKNOWLEDGMENTS

First of all, I would like to gratefully and sincerely thank my advisor Dr. Jiancheng Jiang for his excellent guidance, caring and patience during my graduate studies at UNC Charlotte. Besides providing me with an effective guidance on my research, his enthusiasm to work also inspired me to push myself for better. This dissertation would not be possible without his guidance and encouragement. He and his family have also offered a lot of help during my whole phd life.

Appreciation also extends to Dr. Yanqing Sun, Dr. Weihua Zhou, and Dr. Aixi Zhou for serving as very valuable members in my advisory committee who gave very precious suggestions and great support on my research.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1: INTRODUCTION AND RELATED LITERATURE

It is well known that likelihood ratio theory is one of the most important statistic results and it develops a useful principle that generally applicable to most parametric hypothesis problems. A key fundamental property that contributes very significantly to the success of the maximum likelihood ratio tests is that their asymptotic null distribution are independent of nuisance parameters. This property was referred to as the Wilks phenomenon in Fan, Zhang and Zhang (2001) as well in this paper. Naturally, the questions how such a useful principle can be extended to multi-dimensional problems and whether the Wilks type of results continue to hold arise. Many computationally intensive nonparametric techniques and theories have been rapidly developed to exploit hidden structures and to reduce modelling biases of traditional parametric methods. Methods such as local linear fitting, local polynomial fitting, orthogonal series expansions and spline approximations, also dimensionality reduction techniques have been studied in great details in many statistical contexts. Yet, there are no generally applicable methods available for the inferences in multivariate nonparametric models. Owen(1988) extending the scope of the likelihood inferences through the empirical likelihood which is applicable to a class of nonparametric functionals. Usually, these functionals are smooth that they can be estimated at root-n rate. See also Owen (1990), Hall and Owen (1993),Chen and Qin (1993), Li, Hollander, Mckeague and Yang(1996) for applications of the empirical likelihood. A further extension of the empirical likelihood, called "random-sieve likelihood", can be found in Shen, Shi and Wong(1999). The random-sieve likelihood allows one to handle the situations where observable variables and stochastic errors are not necessarily one-on-one. In addition, various efforts have been put on nonparametric hypothesis testing. For instance, see, Bickel and Ritov (1992),Fan(1996), Fan and Li(1996), Kallenberg and Ledwina (1997). However, most of the studies focus only on the one-dimensional nonparametric regression problem. It is difficult to extend them to multivariate semiparametric and nonparametric models. In order to derive a generally applicable testing procedure for multivariate semiparametric and nonparametric models. Fan, Zhang and Zhang (2001) proposed generalized likelihood ratio tests. The work is motivated by the fact that the nonparametric maximum likelihood ratio test may not exist in many nonparametric problems. Generalized likelihood ratio statistics, obtained by replacing unknown functions by reasonable nonparametric estimators have several nice properties. For instance additive models (Fan and Jiang 2005):

$$\mathbf{Y} = m_1(\mathbf{X}_1) + \dots + m_p(\mathbf{X}_p) + \boldsymbol{\varepsilon} \tag{1.1}$$

or time-varying coefficient models(Dr Hoover 1998):

$$\mathbf{Y} = a_1(t/T)\mathbf{X}_1 + \dots + a_p(t/T)\mathbf{X}_p + \boldsymbol{\varepsilon} \tag{1.2}$$

where $\mathbf{X}_1, \dots, \mathbf{X}_p$ are covariates. One would ask if certain parametric forms such as linear models fit the data adequately, after fitting the model. This means testing if each additive component is linear in the additive model (1.1) or if the coefficient functions in (1.2) are really time-varying or not.

1.1 Generalized likelihood ratios

Let us begin with a simple nonparametric regression model to motivate the generalized likelihood ratio statistics. Suppose we have n sample data $\{X_i, Y_i\}$ from the nonparametric regression model, for $i = 1, ...n$,

$$Y_i = m(X_i) + \varepsilon_i \tag{1.3}$$

where $\{\varepsilon_i\}$ are a sequence of i.i.d. random variables from $N(0, \sigma^2)$ and X_i has a density f with support $[0, 1]$. Denote the parameter space is

$$\mathcal{F}_k = \{m \in L^2[0,1] : \int m^{(k)}(x)^2 dx \leqslant C\} \tag{1.4}$$

for a given C. Consider the testing problem:

$$H_0 : m(x) = \alpha_0 + \alpha_1 x \longleftrightarrow H_1 : m(x) \neq \alpha_0 + \alpha_1 x \tag{1.5}$$

Hence, the conditional log-likelihood function is: $l_n(m) = -nlog(\sqrt{2\pi}\sigma) - \frac{1}{2\sigma^2}\sum_{i=1}^n (Y_i - m(X_i))^2$. Let $(\hat{\alpha}_0, \hat{\alpha}_1)$ be the maximum likelihood estimator (MLE) under H_0, and $\hat{m}_{MLE}(\cdot)$ be the MLE under the full model: $min \sum_{i=1}^n (Y_i - m(X_i))^2$, subject to $\int m^{(k)}(x)^2 dx \leqslant C$. The resulting estimator $\hat{m}_{MLE}$ is a smoothing spline. Define the residual sum of squares RSS_0 and RSS_1 as follows:

$$RSS_0 = \sum_{i=1}^n (Y_i - \hat{\alpha}_0 - \hat{\alpha}_1 X_i)^2, RSS_1 = \sum_{i=1}^n (Y_i - \hat{m}_{MLE}(X_i))^2. \tag{1.6}$$

So it is easy to see that the logarithm of the conditional maximum likelihood ratio statistic for the problem (1.5) is given by:

$$\lambda_n = l_n(\hat{m}_{MLE}) - l_n(H_0) = \frac{n}{2} log \frac{RSS_0}{RSS_1} \approx \frac{n}{2}\frac{RSS_0 - RSS_1}{RSS_1}$$

Technically, the maximum likelihood ratio test is not convenient to manipulate and is either not optimal due to restriction of choosing smoothing parameters. In general, MLEs under nonparametric regression models are hard to obtain. Therefore, we replace the maximum likelihood estimator under the alternative nonparametric model by any reasonable nonparametric estimator, giving the generalized likelihood ratio

$$\lambda_n = l_n(H_1) - l_n(H_0) \tag{1.7}$$

Here $l_n(H_1)$ is the likelihood with unknown regression function replaced by a reasonable nonparametric regression estimator. We can find similar ideas in Severini and Wong (1992) for construction of semi-parametric efficient estimators. We notice that the nonparametric estimator does not have to belong to $\mathcal{F}_k$. Thus the assumption that the constant C in (1.4) is given can be removed. The above generalized likelihood method can be readily be applied to other statistical methods such as additive models, varying-coefficient models, and any nonparametric regression model with a parametric regression model with a parametric error distribution. Using suitable nonparametric estimators, we need to compute the likelihood function under null and alternative models. The generalized likelihood ratio tests are expected to be powerful with appropriate choice of smoothing parameters.

1.2 Wilks phenomenon

Based on the local linear estimators (Fan, 1993), we are going to show the asymptotic null distribution of the generalized likelihood ratio statistic is nearly χ^2 with large degrees of freedom in the sense that

$$\gamma\lambda_n \sim \chi^2_{\mu_n} \tag{1.8}$$

for a sequence $\mu_n \longrightarrow \infty$ and a constant σ, namely, $(2\mu_n)^{-1/2}(\gamma\lambda_n - \mu_n) \longrightarrow N(0,1)$ in distribution. It means that the asymptotic null distribution is independent of the nuisance parameters α_0, α_1 and σ and the design density function f. Since the Wilks phenomenon holds, we can simply simulate the null distributions and then get the constants μ_n and γ without theoretically deriving the values of them to be able to use generalized likelihood ratio test. This newly discovered Wilks phenomenon is the key to the success of the generalized likelihood ratio tests for nonparametric problems. The p-values can be easily be computed from the asymptotic distribution or simulations by fixing nuisance parameters under the null hypothesis. Furthermore, Fan,Zhang and Zhang(2001) has shown that the resulting tests are asymptotic optimal in the sense of Ingster(1993).

1.3 Related literature

There are various collective efforts on hypothesis testing in nonparametric regression models,most of which focus on one dimensional problems. For an overview, see the recent book by Hart(1997). Bickel and Rosenblatt(1973) gave the asymptotic null distributions in their early paper on nonparametric hypothesis testing. A few new nonparametric testing techniques have been proposed in Bickel and Ritov(1992). For the Cox's hazard regression model, Murphy(1993) derived a Wilks type of result for a generalized likelihood ration statistic based on a simple sieve estimator. Fan (1996) proposed simple and powerful methods for constructing tests based on Neyman's truncation and wavelet thresholding. It was proved in Spokoiny(1996) that wavelet thresholding tests are nearly adaptively minimax. Hypothesis testing for multivariate regression problems is not easy because of the curse of dimensionality. In bivariate regression, Aerts *et al.*(1999) built tests based on orthogonal series. Fan and Huang(1998) came up with several testing techniques on the basis of the adaptive Nayman test for many alternative models in multiple regression setting. These problems become conceptually simple by using our generalized likelihood method. But for the multivariate time-varying coefficient nonparametric regression models, to the best of my knowledge, there is no paper in the literature to develop the similar test procedure. In this dissertation,I develop some new test procedure, termed as generalized quasi-likelihood ratio (GQLR) test, to check whether coefficients are of constant or specific functional form. I will be presenting the proposed GQLR test procedure in details in chapter 3 of this paper.

1.4 Outline of the paper

The rest of this dissertation is organized as follows. In Chapter 2, I discuss the estimation of coefficients in a time-varying coefficient multivariate regression model by using local linear technique and then derive the explicit expression of the proposed estimator. Further, I derive the asymptotic theory for the nonparametric estimator.

In Chapter 3, I propose the new so-called GQLR test for the multivariate time-varying coefficient model to test if varying coefficients for the time-varying nonparametric regression model are some known constants or of some specific time-varying functional forms. The test statistics are constructed based on the comparison of the quasi-likelihood under null and alternative hypotheses respectively. I derive the asymptotic distributions of the test statistic under null and alternative hypotheses,also show how the Wilks type of result continue to hold in this multi-dimensional case. In addition, Monte Carlo simulation is done to show the finite sample performance of the proposed methods, power curves are presented for different error distributions and different sample sizes. Finally an real example of monthly US interest rate is given for the application of the methodology.

Chapter 4 concludes the dissertation. The detailed proofs of the main results in each chapter are relegated to the last section of the corresponding chapter.

CHAPTER 2: ESTIMATION AND ASYMPTOTIC RESULTS

This chapter mainly discusses the local linear estimation of the time-varying coefficient and asymptotic distribution of the nonparametric estimator.

2.1 The model

I propose the multivariate time-varying coefficient model:

$$\mathbf{y}_t = \mathbf{c}(t/T) + \sum_{i=1}^{p} \boldsymbol{\alpha}_i(t/T)\mathbf{y}_{t-i} + \sum_{j=1}^{q} \boldsymbol{\beta}_j(t/T)\mathbf{x}_{t-j} + \boldsymbol{\varepsilon}_t, t = 1, ..., T. \tag{2.1}$$

where $\mathbf{y}_t$ is $k \times 1$ vector, $\mathbf{x}_t$ is $v \times 1$ vector. $\mathbf{c}(.)$ is a $k \times 1$ vector, $\boldsymbol{\alpha}_i$ is $k \times k$ smooth matrix and $\boldsymbol{\beta}_j$ is $k \times v$ smooth matrix. The innovations satisfy $\boldsymbol{\varepsilon}_t = \gamma_t^* \mathbf{a}_t$, where γ_t^* are symmetric positive definite matrices and $\mathbf{a}_t$ is a sequence of uncorrelated random vectors with mean zero and convariance matrix $\mathbf{I}_k$. Let $\mathbf{X}_t = vec(1, \mathbf{y}_{t-1}, ..., \mathbf{y}_{t-p}, \mathbf{x}_{t-1}, ..., \mathbf{x}_{t-q})$ be a $d \times 1$ vector with $d = 1 + kp + vq$, and $\boldsymbol{\Phi}(t/T) = (\mathbf{c}(t/T), \boldsymbol{\alpha}_1(t/T), ...\boldsymbol{\alpha}_p(t/T), \boldsymbol{\beta}_1(t/T), ...\boldsymbol{\beta}_q(t/T)$, Then model (2.1) becomes:

$$\mathbf{y}_t = \boldsymbol{\Phi}(t/T)\mathbf{X}_t + \boldsymbol{\varepsilon}_t, t = 1, ...T. \tag{2.2}$$

where $\boldsymbol{\Phi}(.)$ is $k \times d$ matrix and $\mathbf{X}_t$ is $d \times 1$ vector.

For any t in the neighborhood of $t_0 \in (0, T)$, i.e $\mid \frac{t - t_0}{T} \mid \leq h$, using the Taylor expansion, we obtain:

$$\begin{aligned} \boldsymbol{\Phi}(t/T) &\approx \boldsymbol{\Phi}(t_0/T) + \boldsymbol{\Phi}'(t_0/T)(\frac{t - t_0}{T}) \\ &\equiv \mathbf{P} + \mathbf{Q}(\frac{t - t_0}{T}) \end{aligned}$$

Running the local linear smoother for model (2.2), we minimize:

$$\sum_{t=s+1}^{T} \parallel \mathbf{y}_t - \mathbf{P}\mathbf{X}_t - \mathbf{Q}\mathbf{X}_t(\frac{t-t_0}{T}) \parallel^2 K_h(t-t_0) \tag{2.3}$$

over $\mathbf{P}$ and $\mathbf{Q}$, where $\parallel \textbf{.} \parallel$ denotes the Euclidean norm, $s = max(p,q)$ and $K_h(x) = \frac{1}{h}K(\frac{x}{hT})$ for a kernel function $K(\textbf{.})$ and a bandwidth h controlling the amount of smoothing. Let the resulting minimizers for $(\mathbf{P}, \mathbf{Q})$ be $(\hat{\mathbf{P}}, \hat{\mathbf{Q}})$.

Let $\mu_i = \int u^i K(u)du$, $\nu_i = \int u^i K^2(u)du$, $\omega = (\mu_2, \mu_3)^T$,

$$U = \begin{pmatrix} \mu_0 & \mu_1 \\ \mu_1 & \mu_2 \end{pmatrix}, V = \begin{pmatrix} \nu_0 & \nu_1 \\ \nu_1 & \nu_2 \end{pmatrix}$$

2.2 Estimation and Asymptotic distribution

Define $\mathbf{M} = E[(\boldsymbol{X}_t\boldsymbol{X}_t^T) \otimes I_k]$ and $\mathbf{N} = E[(\boldsymbol{X}_t\boldsymbol{X}_t^T) \otimes (\gamma_t^*)^2]$. Next, I derive the explicit representation of the estimator by using local linear fitting .

Theorem 2.1 The solution $(\hat{\mathbf{P}}, \hat{\mathbf{Q}})$ for (2.3) admits the following closed form:

$$\begin{bmatrix} vec(\hat{\mathbf{P}}) \\ vec(h\hat{\mathbf{Q}}) \end{bmatrix} = \begin{pmatrix} \mathbf{S}_{T0} & \mathbf{S}_{T1} \\ \mathbf{S}_{T1} & \mathbf{S}_{T2} \end{pmatrix}^{-1} \begin{bmatrix} \sum_{t=s+1}^{T}(\mathbf{X}_t \otimes I_k)\mathbf{y}_t K_h(t-t_0) \\ \sum_{t=s+1}^{T}(\mathbf{X}_t \otimes I_k)\mathbf{y}_t K_h^{(1)}(t-t_0) \end{bmatrix} \tag{2.4}$$

where $\otimes$ denotes the kronecker product. I_k is the $k \times k$ indentity matrix, $\mathbf{S}_{Ti} = \sum_{t=s+1}^{T}(\mathbf{X}_t\mathbf{X}_t^T) \otimes I_k K_h^{(i)}(t-t_0)$ and $K_h^{(i)}(t-t_0) = (Th)^{(-i)}(t-t_0)^i K_h(t-t_0)$, for $i = 0, 1, 2$.

Proof: See Section 2.3. $\square$.

To derive the asymptotic distribution of the above estimators, we need the following assumptions.

Assumption A:

(A1) For any $u = t_0/T \in (0,1)$, the second derivative of $\mathbf{\Phi}(\mathbf{.})$ exists and is continuous at u.

(A2) The kernel function $K(v)$ is symmetrical with a bounded support s.t $\mu_0(K) = 1$ and $\mu_1(K) = 0$ i.e. $\int K(v)dv = 1$ and $\int vK(v)dv = 0$. Further, the functions $v^3K(v)$ and $v^3K'(v)$ are bounded with $v^4K(v) < \infty$.

(A3) There exists a positive $\rho > 0$ such that $E \parallel \mathbf{a}_t \parallel^{1+\rho} < \infty$.

(A4) Assume that γ_t^* is measurable with respect to the σ-field generated by the historical information $\mathcal{F}_{t-1} = \{\mathbf{w}_s; s \leq t-1\}$, where $\mathbf{w}_{t-1} = (\mathbf{y}_{t-1}, \mathbf{y}_{t-2}, ...\mathbf{y}_{t-p}, \mathbf{x}_{t-1}, ..., \mathbf{x}_{t-q})$.

(A5) $\mathbf{M}$ and $\mathbf{N}$ are both invertible positive definite matrices.

(A6) The processes $\{\mathbf{X}_t, \mathbf{a}_t\}$ are strictly stationary with α-mixing coefficients $\alpha(s)$ such that $\sum_s s^c[\alpha(s)]^{1-2/\delta} < \infty$ for some $\delta > 2$ and $c > 1 - 2/\delta$.

(A7) $h \longrightarrow 0$ in such a way that $hT \longrightarrow 0$. There exists a sequence of positive integers $\{r_T\}$ s.t. $r_T \longrightarrow \infty$, $r_T = o(\sqrt{hT})$ and $\sqrt{T/h}\alpha(r_T) \longrightarrow 0$ as $T \longrightarrow \infty$.

Theorem 2.2 Suppose the assumptions (A1)-(A7) hold. Then, for any $u = t_0/T \in (0,1)$, we have:

$$\sqrt{Th}\{\begin{pmatrix} vec(\hat{\mathbf{P}} - \mathbf{\Phi}(u)) \\ vec(h(\hat{\mathbf{Q}} - \mathbf{\Phi}'(u))) \end{pmatrix} - \mathbf{B}_T(u)\} \xrightarrow{\mathcal{D}} \mathcal{N}(\mathbf{0}, \mathbf{\Sigma})$$

where $\mathbf{B}_T(u) = \dfrac{h^2}{2}(U^{-1}\omega)\otimes vec(\mathbf{\Phi}"(u))(1+o_p(1))$ and $\mathbf{\Sigma} = (U^{-1}VU^{-1})\otimes(\mathbf{M}^{-1}\mathbf{N}\mathbf{M}^{-1})$.

Proof: See Section 2.3. $\square$.

2.3 Complements

In this section, we give the derivation of the main results presented in previous sections of this chapter.

Proof of Theorem 2.1: We can prove it by following the similar steps in the proof of Lemma 1 in Jiang (2013): The proof involves taking the derivative of a generic matrix-valued function $\mathbf{F}(\mathbf{X})$ with respect to a matrix $\mathbf{X}$. Taking derivative over $\mathbf{P}$ and $\mathbf{Q}$ for (2.3), we obtain the score equations:

$$\left\{\begin{array}{c} \sum_{t=s+1}^{T}(\mathbf{X}_t \otimes I_k)[\mathbf{y}_t - \hat{\mathbf{P}}\mathbf{X}_t - \hat{\mathbf{Q}}(\frac{t-t_0}{T})]K_h(t-t_0) = 0 \\ \sum_{t=s+1}^{T}(\mathbf{X}_t \otimes I_k)[\mathbf{y}_t - \hat{\mathbf{P}}\mathbf{X}_t - \hat{\mathbf{Q}}(\frac{t-t_0}{T})]K_h^{(1)}(t-t_0) = 0 \end{array}\right. \tag{2.5}$$

For conforming matrices, we have the identity:

$$vec(\mathbf{AXB}) = (\mathbf{B}^T \otimes \mathbf{A})vec(\mathbf{X}) \tag{2.6}$$

This combined with the identity:

$$(\mathbf{A} \otimes \mathbf{B}) \otimes \mathbf{C} = \mathbf{A} \otimes (\mathbf{B} \otimes \mathbf{C}) \tag{2.7}$$

yields that

$$\begin{aligned} (\mathbf{X}_t \otimes I_k)\hat{\mathbf{P}}\mathbf{X}_t &= vec((\mathbf{X}_t \otimes I_k)\hat{\mathbf{P}}\boldsymbol{X}_t) \\ &= ((\mathbf{X}_t^T \otimes \mathbf{X}_t) \otimes I_k)vec(\hat{\mathbf{P}}) \\ &= ((\mathbf{X}_t\mathbf{X}_t^T) \otimes I_k)vec(\hat{\mathbf{P}}) \end{aligned} \tag{2.8}$$

it follows from (2.5) that:

$$\left\{\begin{array}{c} \mathbf{S}_{T0}vec(\hat{\mathbf{P}}) + \mathbf{S}_{T1}vec(h\hat{\mathbf{Q}}) = \sum_{t=s+1}^{T}(\mathbf{X}_t \otimes I_k)\mathbf{y}_t K_h(t-t_0) \\ \mathbf{S}_{T1}vec(\hat{\mathbf{P}}) + \mathbf{S}_{T2}vec(h\hat{\mathbf{Q}}) = \sum_{t=s+1}^{T}(\mathbf{X}_t \otimes I_k)\mathbf{y}_t K_h^{(1)}(t-t_0) \end{array}\right. \tag{2.9}$$

□.

Proof of Theorem 2.2: By taking iterative expectation, we get that:

$$E(T^{-1}\mathbf{S}_{Ti}) = \frac{1}{T}E(\sum_{t=s+1}^{T}(\mathbf{X}_t\mathbf{X}_t^T) \otimes I_k K_h^{(i)}(t-t_0))$$

$$= \frac{1}{T}\sum_{t=s+1}^{T} E[(\mathbf{X}_t\mathbf{X}_t^T)\otimes I_k]K_h^{(i)}(t-t_0)\frac{\mathbf{M}}{T}\int K_h^{(i)}(t-t_0)dt(1+o(1))$$

$$= \mu_i\mathbf{M}(1+o(1)).$$

Note that $T^{-1}vec(\mathbf{S}_{Ti}) = T^{-1}\sum_{t=1}^{T-s}\mathbf{Z}_t$, where $\mathbf{Z}_t = vec[(\mathbf{X}_{s+t}\mathbf{X}_{s+t}^T)\otimes I_k]K_h^{(i)}(s+t-t_0)$. It follows from stationarity that

$$Var(T^{-1}vec(\mathbf{S}_{Ti})) = \frac{T-s}{T^2}Var(\mathbf{Z}_1) + \frac{2(T-s)}{T^2}\sum_{l=1}^{T-s-1}(1-\frac{l}{T-s})cov(\mathbf{Z}_1,\mathbf{Z}_{l+1}) \tag{2.10}$$

Let $d_T \longrightarrow \infty$ be a sequence of integers such that $d_T h \longrightarrow 0$. Define $J_1 = \sum_{l=1}^{d_T-1} \mid cov(\mathbf{Z}_1,\mathbf{Z}_{l+1}) \mid$ and $J_2 = \sum_{l=d_T}^{T-s-1} \mid cov(\mathbf{Z}_1,\mathbf{Z}_{l+1}) \mid$. Using the mixing condition (A6) and Davydov's lemma (see Hall and Heyde 1980, cor.A.2), we have: for components of $\mathbf{Z}_1$ and $\mathbf{Z}_{l+1}$,

$$\mid cov(\mathbf{Z}_{1,j},\mathbf{Z}_{l+1,m}) \mid \leq C[\alpha(l)]^{1-2/\delta}[E|\mathbf{Z}_{1,j}|^{\delta}]^{1/\delta}[E|\mathbf{Z}_{l+1,m}|^{\delta}]^{1/\delta}.$$

where C is a generic constant. Directly calculating the mean and covariance, we establish that: $E|\mathbf{Z}_1|^{\delta} = O(h^{-\delta+1})$ and $|cov(\mathbf{Z}_1,\mathbf{Z}_{l+1})| = O(1)$, componentwise. Then $J_1 = O(d_T) = o(h^{-1})$ and $J_2 = O(h^{2/\delta-2})\sum_{l=d_T}^{\infty}[\alpha(l)]^{1-2/\delta} = O(h^{2/\delta-2})$ $d_T^{-c}\sum_{l=d_T}^{\infty} l_{-c}[\alpha(l)]^{1-2/\delta} = o(h^{-1})$, if we set $h^{1-2/\delta}d_T^c = 1$, so that $d_T h \longrightarrow 0$ is satisfied. Thus,

$$\sum_{l=1}^{T-s-1}|cov(\mathbf{Z}_1,\mathbf{Z}_{l+1})| = J_1 + J_2 = o(h^{-1}) \tag{2.11}$$

Note that $var(\mathbf{Z}_1) = h^{-1}\nu_{2i}E[vec(\mathbf{X}_t\mathbf{X}_t^T)\otimes I_k]^{\otimes 2}(1+o(1))$. It follows from (2.9) and (2.10) that

$$Var(T^{-1}vec(\mathbf{S}_{Ti})) = \frac{1}{Th}\nu_{2i}E[vec(\mathbf{X}_t\mathbf{X}_t^T)\otimes I_k]^{\otimes 2}(1+o(1)) \tag{2.12}$$

By the Chebyshev inequality, we know:

$$T^{-1}\mathbf{S}_{Ti} = \mu_i \mathbf{M}(1 + o_p(1)) \tag{2.13}$$

Hence,

$$T^{-1}\begin{pmatrix} \mathbf{S}_{T0} & \mathbf{S}_{T1} \\ \mathbf{S}_{T1} & \mathbf{S}_{T2} \end{pmatrix} = U \otimes \mathbf{M}(1 + o_p(1)) \tag{2.14}$$

By (2.3) and (2.4), we have

$$\begin{bmatrix} vec(\hat{\mathbf{P}} - \mathbf{\Phi}(u)) \\ vec(h(\hat{\mathbf{Q}} - \mathbf{\Phi}'(u))) \end{bmatrix} = \mathbf{B}_T(u) + \mathbf{V}_T(u) \tag{2.15}$$

where

$$\mathbf{B}_T(u) = \begin{pmatrix} \mathbf{S}_{T0} & \mathbf{S}_{T1} \\ \mathbf{S}_{T1} & \mathbf{S}_{T2} \end{pmatrix}^{-1} \begin{bmatrix} \sum_{t=s+1}^{T}(\mathbf{X}_t \otimes I_k)\mathbf{\Phi}(t/T)\mathbf{X}_t K_h(t-t_0) \\ \sum_{t=s+1}^{T}(\mathbf{X}_t \otimes I_k)\mathbf{\Phi}(t/T)\mathbf{X}_t K_h^{(1)}(t-t_0) \end{bmatrix} - \begin{bmatrix} vec(\mathbf{\Phi}(u)) \\ vec(h\mathbf{\Phi}'(u)) \end{bmatrix},$$

$$\mathbf{V}_T(u) = \begin{pmatrix} \mathbf{S}_{T0} & \mathbf{S}_{T1} \\ \mathbf{S}_{T1} & \mathbf{S}_{T2} \end{pmatrix}^{-1} \begin{bmatrix} \sum_{t=s+1}^{T}(\mathbf{X}_t \otimes I_k)\boldsymbol{\varepsilon}_t K_h(t-t_0) \\ \sum_{t=s+1}^{T}(\mathbf{X}_t \otimes I_k)\boldsymbol{\varepsilon}_t K_h^{(1)}(t-t_0) \end{bmatrix} \equiv \begin{pmatrix} \mathbf{S}_{T0} & \mathbf{S}_{T1} \\ \mathbf{S}_{T1} & \mathbf{S}_{T2} \end{pmatrix}^{-1} \begin{bmatrix} \mathbf{V}_{T0}^{*} \\ \mathbf{V}_{T1}^{*} \end{bmatrix}.$$

Thus $\mathbf{B}_T(u)$ and $\mathbf{V}_T(u)$ contribute to the bias and variance of the estimators, respectively. By the definition of $\mathbf{S}_{Ti}$ and (2.7), we have:

$$\begin{aligned} \mathbf{S}_{T0} vec(\mathbf{\Phi}(u)) &= \sum_{t=s+1}^{T} (\mathbf{X}_t^T \mathbf{X}_t) \otimes I_k vec(\mathbf{\Phi}(u)) K_h(t-t_0) \\ &= \sum_{t=s+1}^{T} \mathbf{X}_t^T \otimes (\mathbf{X}_t \otimes I_k) vec(\mathbf{\Phi}(u)) K_h(t-t_0) \end{aligned}$$

using (2.6), we obtain:

$$\mathbf{S}_{T0}vec(\mathbf{\Phi}(u)) = \sum_{t=s+1}^{T} vec((\mathbf{X}_t \otimes I_k)\mathbf{\Phi}(u)\mathbf{X}_t)K_h(t-t_0)$$
$$= \sum_{t=s+1}^{T} (\mathbf{X}_t \otimes I_k)\mathbf{\Phi}(u)\mathbf{X}_t K_h(t-t_0)$$

Similarly,

$\mathbf{S}_{T1}vec(\mathbf{\Phi}(u) = \sum_{t=s+1}^{T} vec((\mathbf{X}_t \otimes I_k)\mathbf{\Phi}(u)\mathbf{X}_t)K_h^{(1)}(t-t_0)$

$\mathbf{S}_{T1}vec(h\mathbf{\Phi}'(u) = \sum_{t=s+1}^{T} vec((\mathbf{X}_t \otimes I_k)h\mathbf{\Phi}'(u)\mathbf{X}_t)K_h^{(1)}(t-t_0)$

$\mathbf{S}_{T2}vec(h\mathbf{\Phi}'(u) = \sum_{t=s+1}^{T} vec((\mathbf{X}_t \otimes I_k)h\mathbf{\Phi}'(u)\mathbf{X}_t)K_h^{(2)}(t-t_0)$

which gives that:

$$\begin{pmatrix} \mathbf{S}_{T0} & \mathbf{S}_{T1} \\ \mathbf{S}_{T1} & \mathbf{S}_{T2} \end{pmatrix} \mathbf{B}_T(u) \equiv \begin{pmatrix} \mathbf{B}_{T0}^* \\ \mathbf{B}_{T1}^* \end{pmatrix} \tag{2.16}$$

where $\mathbf{B}_{Tj}^* = \sum_{t=s+1}^{T}(\mathbf{X}_t \otimes I_k)[\mathbf{\Phi}(t/T) - \mathbf{\Phi}(u) - \mathbf{\Phi}'(u)\frac{(t-t_0)}{T}]\mathbf{X}_t K_h^{(j)}(t-t_0)$, for $j = 0, 1$. By the Taylor expansion, it can be shown that:

$$T^{-1}\mathbf{B}_{T0}^* = \frac{h^2}{2T}\sum_{t=s+1}^{T}(\mathbf{X}_t \otimes I_k)\mathbf{\Phi}"(u)\mathbf{X}_t K_h^{(2)}(t-t_0) + o_p(h^2)$$
$$= \frac{h^2}{2T}\sum_{t=s+1}^{T}[(\mathbf{X}_t^T \otimes \mathbf{X}_t) \otimes I_k]vec(\mathbf{\Phi}"(u))K_h^{(2)}(t-t_0) + o_p(h^2)$$
$$= \frac{h^2}{2}\mu_2\mathbf{M}vec(\mathbf{\Phi}"(u)) + o_p(h^2)$$

Similarly, $T^{-1}\mathbf{B}_{T1}^* = \frac{h^2}{2}\mu_3\mathbf{M}vec(\mathbf{\Phi}"(u)) + o_p(h^2)$. This combined with (2.8) leads to:

$$\mathbf{B}_T(u) = \frac{h^2}{2}(U^{-1}\otimes\mathbf{M}^{-1})[\omega\otimes(\mathbf{M}vec(\mathbf{\Phi}"(u)))](1+o_p(1))\frac{h^2}{2}(U^{-1}\omega)\otimes vec(\mathbf{\Phi}"(u))(1+o_p(1)) \tag{2.17}$$

where the second equation comes from the identity:

$$(\mathbf{A}\otimes\mathbf{B})(\mathbf{C}\otimes\mathbf{D}) = (\mathbf{AC})\otimes(\mathbf{BD}) \tag{2.18}$$

Let $\mathbf{V}_T^* = (\mathbf{V}_{T0}^{T*}, \mathbf{V}_{T1}^{T*})^T$. Using the same argument as for (2.11), we obtain:

$$Var(\sqrt{T^{-1}h}\mathbf{V}_T^*) = \mathbf{V}\otimes E[(\mathbf{X}_t\otimes I_k)\gamma_t^{*2}(\mathbf{X}_t\otimes I_k)^T](1+o(1)) \tag{2.19}$$

For any unit vector $\mathbf{d} = (\mathbf{d}_1^T, \mathbf{d}_2^T)^T \in \mathcal{R}^{2d}$, where $\mathbf{d}_1$ and $\mathbf{d}_2$ are $d\times 1$ vectors, we get:

$$\begin{aligned}\sqrt{h/T}\mathbf{d}^T\mathbf{V}_T^* &= \frac{1}{\sqrt{T}}\sum_{t=1}^{T-s}\sqrt{h}[\mathbf{d}_1^T(\mathbf{X}_{s+t}\otimes I_k)K_h(t+s-t_0)\\ &+ \mathbf{d}_2^T(\mathbf{X}_{s+t}\otimes I_k)K_h^{(1)}(t+s-t_0)]\boldsymbol{\varepsilon}_{s+t}\\ &\equiv \frac{1}{\sqrt{T}}\sum_{t=1}^{T-s}\mathbf{R}_{T,t}\end{aligned}$$

By (2.18), we get: $Var(\sqrt{h/T}\mathbf{d}^T\mathbf{V}_T^*) = \mathbf{d}^T[\mathbf{V}\otimes\mathbf{N}]\mathbf{d}(1+o(1)) \equiv \boldsymbol{\theta}^2(1+o(1))$. where $\mathbf{N}$ is defined earlier. Applying the "big-block" and "small-block" argument (see the proof of Theorem 6.3, Fan and Yao 2003), we have: $\sqrt{h/T}\mathbf{d}^T\mathbf{V}_T^* \xrightarrow{\mathcal{D}} \mathcal{N}(\mathbf{0}, \boldsymbol{\theta}^2)$. Therefore, by (2.13) and the *Cramér−Wold* device, $\sqrt{Th}\mathbf{V}_T(u)$ is asymptotically normal with mean zero and variance-covariance matrix $\mathbf{\Sigma} = (U\otimes\mathbf{M})^{-1}(V\otimes\mathbf{N})(U\otimes\mathbf{M})^{-1}$, using the identity (2.17), we have: $\mathbf{\Sigma} = (U^{-1}VU^{-1})\otimes(\mathbf{M}^{-1}\mathbf{N}\mathbf{M}^{-1})$. Then the result of the theorem follows. $\square$.

CHAPTER 3: GENERALIZED QUASI-LIKELIHOOD RATIO TEST FOR MULTIVARIATE TIME-VARYING COEFFICIENT REGRESSION MODEL

This chapter briefly discussed the testing hypotheses about whether the coefficients of the time-varying regression models are of some specific functional forms or constants. The simulation studies and a real example application of the proposed test procedure are also presented at the end of this chapter.

3.1 Introduction and Motivation

The likelihood ratio type test was proposed by Fan, Zhang and Zhang (2001) and studied extensively by Fan and Jiang (2005). For the varying coefficient regression model: $Y = A(U)^T\mathbf{X} + \varepsilon,\ \varepsilon \sim \mathcal{N}(0, \sigma^2)$ with $\mathbf{X} = (x_1, ..., xp)^T$ and $A(U) = (a_1(U), ...a_p)^T$. After fitting the regression models via local linear technique. Fan, Zhang and Zhang(2001) raises one interesting problem to check whether the varying coefficients are of some specific functional forms. This is equivalent to the following hypotheses:

$$H_0 : A(U) = A_0(U) \longleftrightarrow H_a : A(U) \neq A_0(U) \tag{3.1}$$

where $A_0(U)$ is a vector of known functionals. One special case of (3.1) is when $A_0(U)$ is a vector of constants. Then the test hypothesis becomes to checking whether the varying coefficients are indeed varying. That is equivalent to:

$$H_0 : A(U) = A_0 \longleftrightarrow H_a : A(U) \neq A_0 \tag{3.2}$$

where A_0 is a vector of known or unknown constants.

The test statistic is defined as:

$$\lambda_n = \mathcal{L}_n(H_a) - \mathcal{L}_n(H_0) = \frac{n}{2}log(\frac{RSS_0}{RSS_a}) \approx \frac{n}{2}(\frac{RSS_0 - RRS_a}{RSS_a})$$

where $\mathcal{L}_n(H_a)$ and $\mathcal{L}_n(H_0)$ are the log-likelihood under H_a and H_0, respectively. $RSS_a = \sum_{k=1}^{n}(Y_k - \hat{A}^T(U_k)\mathbf{X}_k)^2$ and $RSS_0 = \sum_{k=1}^{n}(Y_k - \hat{A}_0^T(U_k)\mathbf{X}_k)^2$. Here $\hat{A}(U)$ is the corresponding nonparametric estimator of $A(U)$ and $\hat{A}_0(U)$ is the true or estimated value of coefficients under H_0. Motivated by Fan, Zhang and Zhang (2001) and Fan and Jiang(2005). Suppose $\{\mathbf{y}_t, \mathbf{x}_t\}_{t=1}^{T}$ are a random sample from the multivariate time-varying coefficient model (2.2). Namely,

$$\mathbf{y}_t = \mathbf{\Phi}(t/T)\mathbf{X}_t + \boldsymbol{\varepsilon}_t \tag{3.3}$$

Now, we assume $\mathbf{\Sigma}_0^{-1/2}\boldsymbol{\varepsilon}_t$ has mean zero and covariance $\mathbf{I}_k$. where $\mathbf{\Sigma}_0$ is a symmetric positive definite constant matrix. $\mathbf{X}_t = vec(1, \mathbf{y}_{t-1}, ..., \mathbf{y}_{t-p}, \mathbf{x}_{t-1}, ..., \mathbf{x}_{t-q})$ is a $d \times 1$ vector with $d = 1 + kp + vq$ and $\mathbf{\Phi}(t/T) = (\mathbf{c}(t/T), \boldsymbol{\alpha}_1(t/T), ...\boldsymbol{\alpha}_p(t/T), \boldsymbol{\beta}_1(t/T), ...$ $\boldsymbol{\beta}_q(t/T))$ is a $k \times d$ matrix.

I consider the simple null hypothesis testing problem:

$$H_0 : \mathbf{\Phi}(t/T) \in \Theta_0(t/T) \longleftrightarrow H_a : \mathbf{\Phi}(t/T) \notin \Theta_0(t/T) \tag{3.4}$$

where $\Theta_0(t/T)$ is a set of functionals of matrix . Denote $\hat{\mathbf{\Phi}}(t/T)$ as the corresponding nonparametric estimator of $\mathbf{\Phi}$. $\hat{\mathbf{\Phi}}_0(t/T)$ is the true or estimated value of coefficients under H_0. I propose the similar test statistic for the testing problem in (3.4) as:

$$\lambda_T = \mathcal{L}(H_a) - \mathcal{L}(H_0) = \frac{T}{2}log(\frac{R\tilde{S}S_0}{R\tilde{S}S_a}) \approx \frac{T}{2}(\frac{R\tilde{S}S_0 - R\tilde{S}S_a}{R\tilde{S}S_a})$$

where $\mathcal{L}(H_a)$ and $\mathcal{L}(H_(0))$ are the log-likelihood under H_a and H_0, respectively. $R\tilde{S}S_a = \sum_{t=1}^{T}(\mathbf{y}_t - \hat{\mathbf{\Phi}}(t/T)\mathbf{X}_t)^T\mathbf{\Sigma}^{-1}(\mathbf{y}_t - \hat{\mathbf{\Phi}}(t/T)\mathbf{X}_t)$ and $R\tilde{S}S_0 = \sum_{t=1}^{T}(\mathbf{y}_t - \hat{\mathbf{\Phi}}_0(t/T)\mathbf{X}_t)^T$

$\mathbf{\Sigma}^{-1}(\mathbf{y}_t - \hat{\mathbf{\Phi}}_0(t/T)\mathbf{X}_t)$ with Σ is a known constant covariance matrix of $\boldsymbol{\varepsilon}_t$ from a working model $\mathbf{y}_t = \mathbf{\Phi}(t/T)\mathbf{X}_t + \boldsymbol{\varepsilon}_t$. Here $\mathbf{\Sigma}$ is invertible and positive definite. So $\mathbf{\Sigma}^{-1}$ can be written in terms of spectral decomposition as:

$$\mathbf{\Sigma}^{-1} = \mathbf{Q}^T \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_k \end{pmatrix} \mathbf{Q}$$

where $\lambda_1 \geqslant \lambda_2 \cdots \geqslant \lambda_k > 0$ are the eigenvalues of $\mathbf{\Sigma}^{-1}$ and $\mathbf{Q}$ is the orthogonal matrix having rows $\mathbf{q}_1, ...\mathbf{q}_k$ which are normalized eigen-vectors corresponding to $\lambda_1, ..., \lambda_k$. With this spectral decomposition of $\mathbf{\Sigma}^{-1}$, model (3.3) becomes equivalently to:

$$\mathbf{Q}\mathbf{y}_t = \mathbf{Q}\mathbf{\Phi}(t/T)\mathbf{X}_t + \mathbf{Q}\boldsymbol{\varepsilon}_t$$

Denote $\mathbf{y}_t^* = \mathbf{Q}\mathbf{y}_t$, $\mathbf{\Phi}^*(t/T) = \mathbf{Q}\mathbf{\Phi}(t/T)$, $\boldsymbol{\varepsilon}_t^* = \mathbf{Q}\boldsymbol{\varepsilon}_t$. Hence, from now on, we focus on the model:

$$\mathbf{y}_t^* = \mathbf{\Phi}^*(t/T) + \boldsymbol{\varepsilon}_t^* \tag{3.5}$$

where $\boldsymbol{\varepsilon}_t^*$ has mean zero and covariance matrix $\mathbf{Q}\mathbf{\Sigma}_0\mathbf{Q}^T$.

Accordingly, the testing hypothesis problem (3.4) becomes:

$$H_0 : \mathbf{\Phi}^*(t/T) \in \Theta_0^*(t/T) \longleftrightarrow H_a : \mathbf{\Phi}^*(t/T) \notin \Theta_0^*(t/T) \tag{3.6}$$

3.2 Test Statistics and Asymptotic Distribution

3.2.1 Test of Functional Form of Time-Varying Coefficients

To derive the asymptotic distribution of $\lambda_T(\mathbf{\Phi}_0^*)$ under H_0, we need the following assumptions.

Assumption B

(B1) $\mathbf{\Phi}^*(u)$ has the continuous second derivative at any $u = t_0/T \in (0, 1)$.

(B2) The kernel function $K(v)$ is symmetrical with a bounded support s.t $\mu_0(K) = 1$ and $\mu_1(K) = 0$ i.e. $\int K(v)dv = 1$ and $\int vK(v)dv = 0$. Further, the functions $v^3K(v)$ and $v^3K'(v)$ are bounded with $v^4K(v) < \infty$.

(B3) $E|\boldsymbol{\varepsilon}_t^*|^4 < \infty$.

(B4) $\mathbf{X}_t$ is bounded. The $d \times d$ matrix $E[\mathbf{X}_t\mathbf{X}_t^T]$ is invertible. $(E[\mathbf{X}_t\mathbf{X}_t^T])^{-1}$ and $E[(\mathbf{X}_t\mathbf{X}_t^T) \otimes \mathbf{\Sigma}_0]$ are both Lipschitz continuous.

We define: $\mathbf{\Gamma} = E[\mathbf{X}_t\mathbf{X}_t^T]$ and $\omega_0 = \iint t^2(s+t)^2K(t/T)K((s+t)/T)dtds$.

Denote $D = k \times d$, $\mathbf{\Omega} \equiv \mathbf{Q\Sigma}_0\mathbf{Q}^T = (\sigma_{ij}^2)_{i,j=1}^k$ i.e, it has (i, j)-th element as σ_{ij}^2, $i, j = 1, ...k$. For $j = 1, 2...k$, Let

$\varepsilon_{tj}^* = y_{tj}^* - \mathbf{\Phi}_{0j}^*(t/T)\mathbf{X}_t$,

$R_{T10}^j = \frac{1}{\sqrt{T}}[\sum_{t=1}^T \varepsilon_{tj}^* \mathbf{\Phi}_{0j}^{"*}\mathbf{X}_t \int t^2K(t/T)dt](1 + O(h) + O(T^{-1/2}))$,

$R_{T20}^j = \frac{1}{2\sqrt{T}} \sum_{t=1}^T \varepsilon_{tj}^* \mathbf{X}_t^T\mathbf{\Gamma}^{-1}(\mathbf{\Phi}_{0j}^{*"}(t/T)\mathbf{X}_t)E(\mathbf{X}_t)\omega_0$,

$R_{T30}^j = \frac{1}{8}E[\mathbf{\Phi}_{0j}^{*"}(t/T)\mathbf{X}_t\mathbf{X}_t^T\mathbf{\Phi}_{0j}^{*"}(t/T)^T]\omega_0(1 + O(T^{-1/2}))$,

$d_{1Tj} = \sigma_{jj}^{-2}[Th^4R_{T30}^j - T^{1/2}h^2(R_{T10}^j - R_{T20}^j)] = O_p(Th^4 + \sqrt{T}h^2)$,

$\mu_T = \frac{D}{h}(K(0) - \frac{1}{2}\int K^2(x)dx)$,

$\sigma_T^2 = \frac{D}{2h}\int (2K(x) - K * K(x))^2dx$,

$d_{1T}^* = \frac{1}{k}\sum_{j=1}^k \lambda_j\sigma_{jj}^2 d_{1Tj}$,

$\mu_T^* = \frac{\mu_T}{k}\sum_{j=1}^k \lambda_j\sigma_{jj}^2$,

$\sigma^{*2} = \frac{\sigma_T^2}{k^2}(\sum_{j=1}^k \lambda_j^2\sigma_{jj}^4 + 2\sum_{i<j}^k \lambda_i\lambda_j\sigma_{ij}^4)$,

where $K * K$ denotes the convolution product of K, note that both R^j_{T10} and R^j_{T20} are asymptotically normal and hence are stochastically bounded.

Then, we have the following theorem.

Theorem 3.1 Suppose Assumption B holds. Then under H_0, as $h \longrightarrow 0$ and $Th^{3/2} \longrightarrow \infty$,

$$\sigma^{*-1}(\lambda_T(\mathbf{\Phi}_0) - \mu^*_T - d^*_{1T}) \xrightarrow{\mathcal{D}} \mathcal{N}(0,1).$$

Proof: See Section 3.6. □

3.2.2 Test of Constancyof Time-Varying Coefficients

One special case of the hypothesis in (3.6) is to check whether the coefficient functions are actually varying. This means when $\Theta^*_0(t/T)$ is some known constant matrix $\mathbf{\Phi}^*_0$. In this case, we have the following asymptotic result.

Theorem 3.2 Suppose Assumption B holds. Then under H_0, as $h \longrightarrow 0$ and $Th^{3/2} \longrightarrow \infty$,

$$\sigma^{*-1}(\lambda_T(\mathbf{\Phi}_0) - \mu^*_T) \xrightarrow{\mathcal{D}} \mathcal{N}(0,1)$$

Proof: See Section 3.6. □

3.3 Power of Test

In this section, we consider the power of the quasi-likelihood ratio test based on local linear fitting. For simplicity, we fix the null hypothesis in (3.6) with a known matrix.

For any $u = t/T \in (0,1)$, if we rewrite matrix $\mathbf{\Phi}^*(u)$ as a vector:

$$\mathbf{\Delta}(u) \equiv vec(\Phi_1^*(u), \Phi_2^*(u), ...\Phi_D^*(u))$$

Denote: $\mathbf{\Delta}_0(u) \equiv vec(\Phi_{01}^*(u), \Phi_{02}^*(u), ...\Phi_{0D}^*(u))$, then the power of the test is considered under the local alternatives as follows:

$$H_a : \mathbf{\Delta}(u) = \mathbf{\Delta}_0(u) + \frac{1}{\sqrt{Th}}\mathbf{G}(u)$$

where $\mathbf{G}(u) = (g_1(u), g_2(u), ..., g_D(u))^T$ is a $D \times 1$ vector of functions. So, the power of the test under H_a can be approximated by using the following theorem.

Theorem 3.3 Suppose that Assumption B holds and $\mathbf{\Delta}(u)$ is linear in u or $Th^5 \longrightarrow 0$. If

$ThE\{\mathbf{G}^T(u)[(\mathbf{X}_t\mathbf{X}_t^T) \otimes I_k]\mathbf{G}(u)\} \longrightarrow C(\mathbf{G})$ and $E\{(\mathbf{G}^T(u)[(\mathbf{X}_t\mathbf{X}_t^T) \otimes I_k]\mathbf{G}(u))$ $\boldsymbol{\varepsilon}_t^{T*}\boldsymbol{\varepsilon}_t^*\}^2 = O((Th)^{-3/2})$ for some constant $C(\mathbf{G})$, then under H_a:

$$\sigma_1^{*-1}(\lambda_T(\mathbf{\Phi}_0) - \mu_T^* - d_{2T}^* + \nu_T^*) \xrightarrow{\mathcal{D}} \mathcal{N}(0,1)$$

where $\sigma_1^{2*} = \sigma^{2*} + \frac{T}{k^2}E\{\mathbf{G}^T(u)[(\mathbf{X}_t\mathbf{X}_t^T) \otimes I_k]\mathbf{G}(u)\}(\sum_{j=1}^k \lambda_j^2\sigma_{jj}^2 + 2\sum_{i<j}^k \lambda_i\lambda_j\sigma_{ij}^2)$,
$d_{2T}^* = \frac{T}{2k}E\{\mathbf{G}^T(u)[(\mathbf{X}_t\mathbf{X}_t^T) \otimes I_k]\mathbf{G}(u)\}(\sum_{j=1}^k \lambda_j\sigma_{jj}^2)$,
$\nu_T^* = \frac{Th^4}{8k}E\{\mathbf{\Delta}^{"T}(u)[(\mathbf{X}_t\mathbf{X}_t^T) \otimes I_k]\mathbf{\Delta}^{"}(u)\}\omega_0(\sum_{j=1}^k \lambda_j)$ and μ_T^* is given in Theorem 3.1.

Proof: See Section 3.6. □.

Remark 3.1: From Theorem 3.3,it is easy to see that the test statistic diverges if $\mathbf{G}(\cdot)$ departs from zero. This implies that the test is consistent.

3.4 Empirical Examples

In this section, I conduct Monte Carlo simulation to demonstrate the power of the proposed QGLR. The effect of the error distribution on the performance of the proposed test is also investigated. Throughout this section, the Gaussian Kernel $K(u) = \frac{1}{\sqrt{2\pi}} e^{-u^2/2}$ is used. Simulation procedures and results are given below.

3.4.1 Simulation Procedures: Conditional Bootstrap

To implement the QGLR tests, we need to obtain the null distribution of the test statistic. In Section 3.2, we give the theoretical asymptotic distribution of the statistic. For a finite sample, the null distribution can be approximated by simulation via fixing nuisance parameters/functions at their reasonable estimates. This simulation method is referred to as the conditional bootstrap, which is detailed as follows:

1) Fix the optimal bandwidth at its estimated value $\hat{h}_{opt}$ and then obtain the estimators of the coefficient $\hat{\mathbf{\Phi}}(t/T)$ under both null and alternative models.

2) Compute the QGLR test statistic $\lambda_T(H_0)$ by definition and the residuals $\mathbf{e}_t$ (for $t = 1, 2, ...T$) from the unrestricted model under H_a.

3) For each X_t, draw a bootstrap residual $\mathbf{e}_t^*$ from the centered empirical distribution of $\mathbf{e}_t$ and compute $\mathbf{y}_t^* = \hat{\mathbf{\Phi}}(t/T)\mathbf{X}_t + \mathbf{e}_t^*$. where $\hat{\mathbf{\Phi}}(t/T)$ is the estimated regression coefficients under H_a in step 1). This forms a conditional bootstrap sample $\{\mathbf{X}_t, \mathbf{y}_t^*\}_{t=1}^T$.

4) Using the bootstrap sample in step 3) with the bandwidth $\hat{h}_{opt}$, obtain the QGLR $\lambda_T^*(H_0)$ in the same manner as $\lambda_T(H_0)$.

5) Repeat steps 3) and 4) many times, say 1000 times to get a sample of statistic $\lambda_T^*(H_0)$. The critical value at significant level α is given by the $(1 - \alpha)$th quantile.

3.4.2 Simulation Results

In this section, I consider the following data generating model:

$$\mathbf{y}_t = \mathbf{\Phi}(t/T)\mathbf{X}_t + \mathbf{e}_t, t = 1, ...T. \tag{3.7}$$

where $k = 2$, $v = p = q = 1$, $D = k \times d = 6$, $\mathbf{\Delta} = vec(\Phi_1, .., \Phi_6) = (0.5, 0.0075, 0.08, 0.65, 0.25, 0.75)^T$, set the initial value $\mathbf{y}_1 = (0.15, 0.2)$ and $x_1 = 0$. In this case, $\mathbf{X}_t = vec(y_{1,t-1}, y_{2,t-1}, x_{t-1})$, for $t = 2, ...T$. $\mathbf{e}_t \sim \mathcal{N}(\mathbf{0}, \Sigma)$. where $\Sigma = \left(\begin{smallmatrix} 1 & 0.5 \\ 0.5 & 1 \end{smallmatrix}\right)$. For the power assessment, we evaluate the power for a sequence of alternative models indexed by θ:

$$H_\theta : \mathbf{\Delta}_\theta = (0.5, 0.0075, 0.08, 0.65, 0.25, 0.75)^T + \frac{\theta}{\sqrt{Th}}\mathbf{G}(t/T) \tag{3.8}$$

where $\mathbf{G}(t/T) = (sin(\sqrt{2}\pi t/T), -0.09cos(\pi t/T), 0.16sin(\sqrt{3}\pi t/T), 0.8sin(\sqrt{2\pi}t/T), 0.3sin(\pi t/T), cos(\sqrt{1.5}\pi t/T))^T$. The simulation is repeated 600 times for each sample size $n = 200$, $n = 400$ and $n = 800$ and for each $\theta = 0$, $\theta = 0.2$, $\theta = 0.4$, $\theta = 0.6$ $\theta = 0.8$ and $\theta = 1.0$. For each given value of θ, I use 1000 Monte Carlo replications for the calculation of the critical values via the conditional bootstrap method (see section 3.4.1). Given the significance of level 5% and 10%, the power function $\rho(\theta)$ is estimated based on the relative frequency of $\lambda_T(\Phi)$ over 600 simulations. In addition to the bivariate normal distribution, the bivariate $t(5)$ and bivariate lognormal$(\mathbf{0}, \Sigma)$ distribution.where Σ is the same variance matrix as of the bivariate normal distribution. I plot the power curves in Figure 3.1 and Figure 3.2 at significance levels 10% and 5% for all settings.

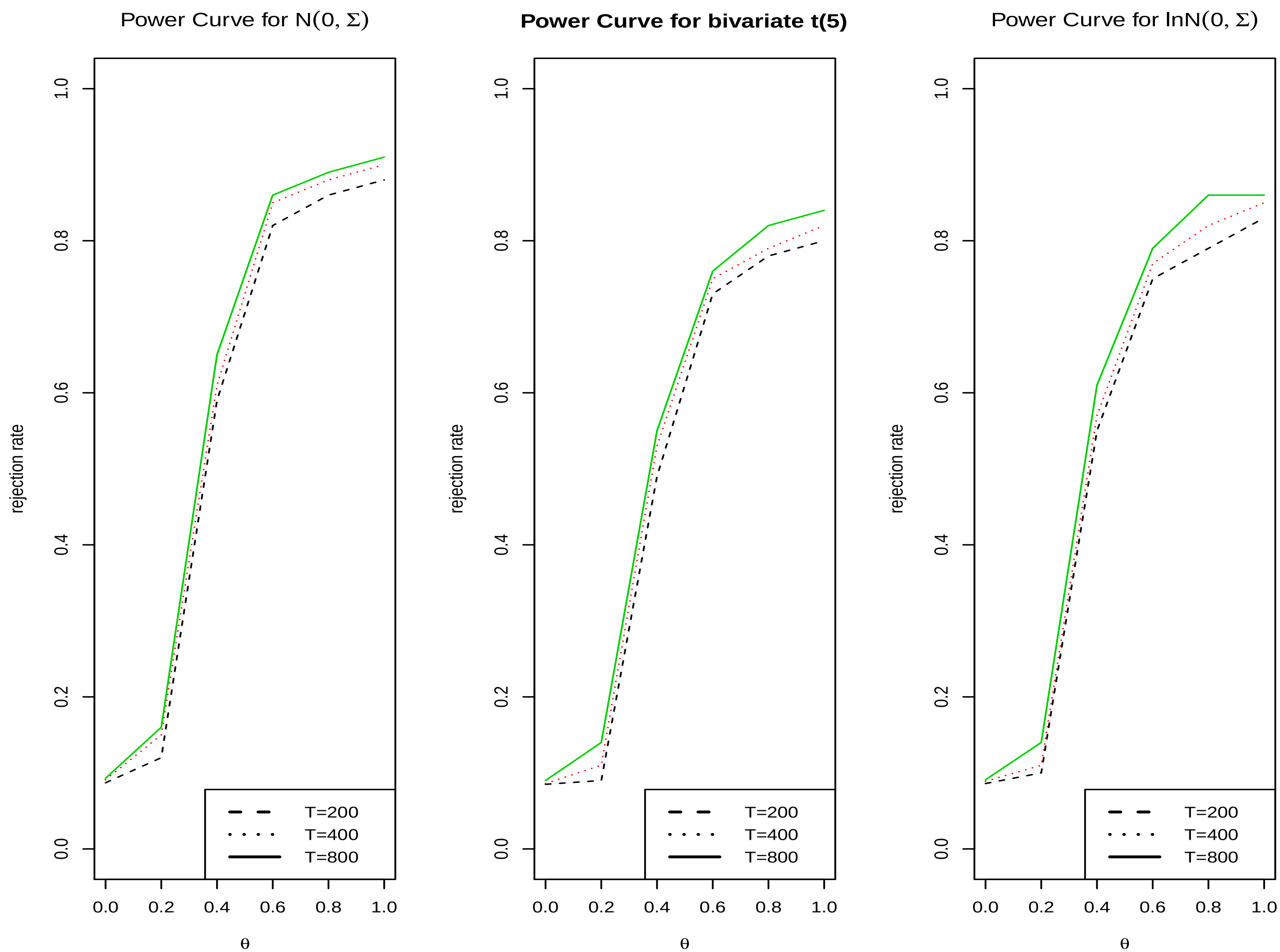

Figure 3.1: The power curves for the testing hypothesis in (3.6) with the nominal size 10%. The dashed line is for $T = 200$, the dotted line is for $T = 400$ and the solid line is for $T = 800$ in Section 3.4.2.

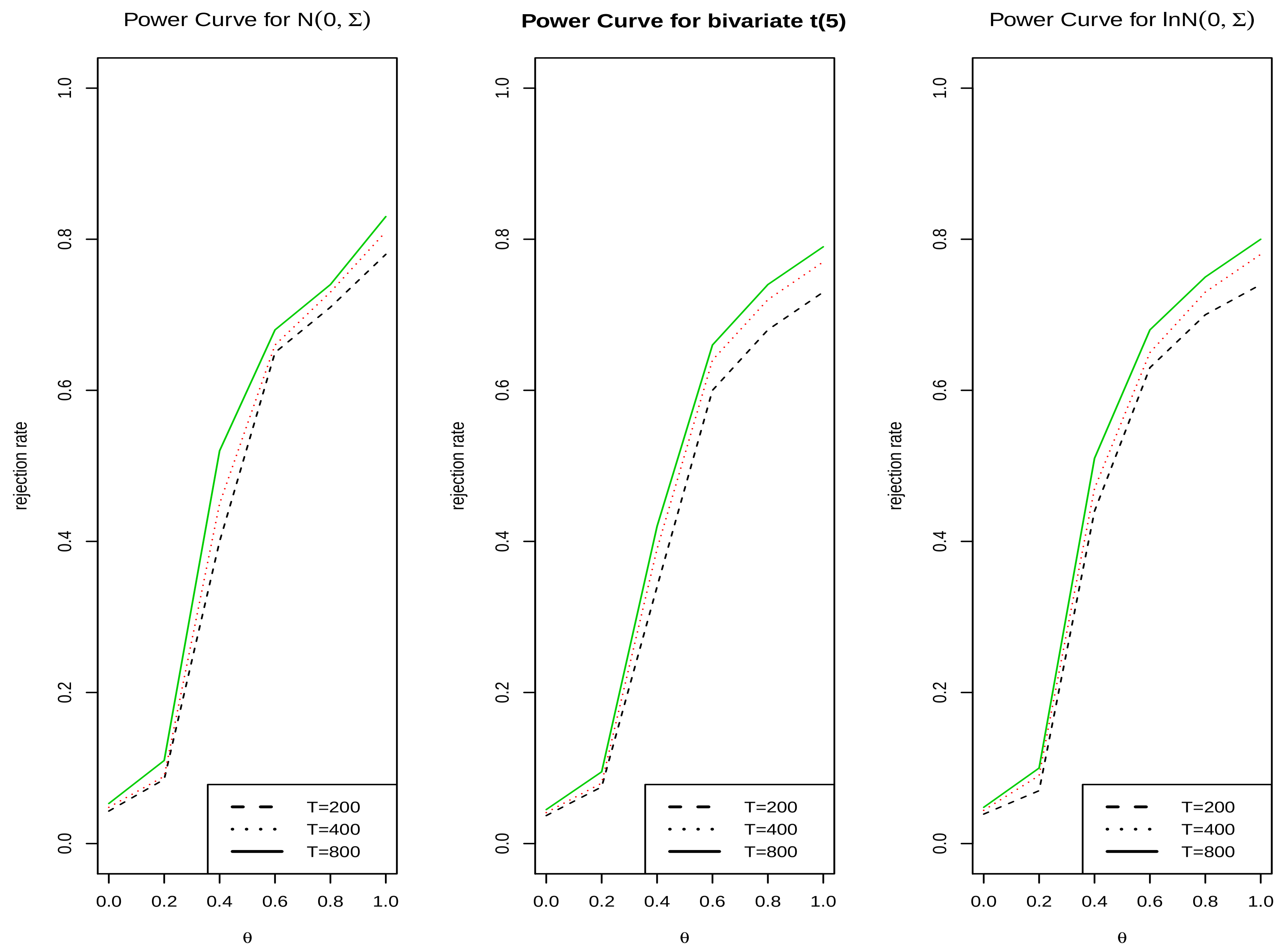

Figure 3.2: The power curves for the testing hypothesis in (3.6) with the nominal size 5%. The dashed line is for $T = 200$, the dotted line is for $T = 400$ and the solid line is for $T = 800$ in Section 3.4.2.

3.5 Eeal Example

In previous section, I conducted Monte Carlo simulation to illustrate the effectiveness and the validity of the proposed test statistics. In this section, I consider the application of these methodologies to a real example. Here I analyze a subset of the interest rates of the Federal Reserve Bank of St.Louis (http://research.stlouisfed.org /fred2/). They are monthly 1-year and 10-year Treasury constant maturity rates,which represent short-term and long-term series, respectively. The data consist of 571 monthly observations from January 1984 to October 2000.

Let Y_{1t} and Y_{2t} be the interest rate series of the 1-year and 10-year Treasury,respectively. Denote $s_{1t} = ln(Y_{1t})$ and $s_{2t} = ln(Y_{2t})$. I use the logarithm return $\mathbf{y}_t = (y_{1t}, y_{2t})^T$, where $y_{it} = s_{it} - s_{i,t-1}, i = 1, 2$. I fit the data using the following bivariate AR(2) model:

$$\mathbf{y}_t = \mathbf{a}\mathbf{y}_{t-1} + \mathbf{b}\mathbf{y}_{t-2} + \boldsymbol{\varepsilon}_t \tag{3.9}$$

where $\mathbf{a} = \left(\begin{smallmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{smallmatrix}\right)$ and $\mathbf{b} = \left(\begin{smallmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{smallmatrix}\right)$. If we rewrite the coefficient matrices into a vector, I now use the proposed GQLR statistic to test the time-varying coefficients $\boldsymbol{\Delta}_\theta = (\hat{a}_{11}, \hat{a}_{21}, \hat{a}_{12}, \hat{a}_{22}, \hat{b}_{11}, \hat{b}_{21}, \hat{b}_{12}, \hat{b}_{22})^T + \dfrac{\theta}{\sqrt{Th}}\mathbf{G}(t/T)$ are indeed time varying. That is to test:

$$H_{0\theta} : \boldsymbol{\Delta}_\theta = (\hat{a}_{11}, \hat{a}_{21}, \hat{a}_{12}, \hat{a}_{22}, \hat{b}_{11}, \hat{b}_{21}, \hat{b}_{12}, \hat{b}_{22})^T + \frac{\theta}{\sqrt{Th}}\mathbf{G}(t/T) \tag{3.10}$$

where $(\hat{a}_{11}, \hat{a}_{21}, \hat{a}_{12}, \hat{a}_{22}, \hat{b}_{11}, \hat{b}_{21}, \hat{b}_{12}, \hat{b}_{22})^T = (0.230, -0.032, 0.334, 0.398, 0.024, 0.008, -0.184, -0.152)^T$ is the estimated coefficients using software R for model (3.9). $\mathbf{G}(t/T) = (0.08sin(\sqrt{2}\pi t/T), 0.3sin(\pi t/T), 0.16sin(\sqrt{3}\pi t/T), cos(\sqrt{1.5}\pi t/T), -0.09 cos(\pi t/T), 0.3sin(\pi t/T), 0.8sin(\sqrt{2\pi}t/T), cos(\sqrt{1.5}\pi t/T))^T$. To compute the p-value of the test statistic, I need to find the null distribution of the GQLR statistic $\lambda_T(H_0)$. This can be estimated by the conditional bootstrap method mentioned in section 3.4.1. The p-values are computed using the optimal bandwidth and from 500 bootstrap replicates for each given θ. The corresponding p values are reported in Table 3.1. Therefore, one can see all the p-values are greater than significant level 0.05 except for $\theta = 0$, which implies that the varying coefficients are indeed time-varying.

Table 3.1: The p-values for testing constancy in model (3.9)

θ	0.0	0.2	0.4	0.6	0.8	1.0
$p - value$	0.032	0.09	0.33	0.21	0.06	0.19

3.6 Complements

In this section, I give the derivations of the main results presented in previous sections of this chapter. Before moving forward to the detailed proofs, we need the following definitions and lemmas.

Let $h_T = 1/\sqrt{Th}$, for each $j = 1, 2..., k$,

$\boldsymbol{\beta}_j(t_0)^T = (\boldsymbol{\Phi}_j^*(t_0/T), h\boldsymbol{\Phi}^{*'}(t_0/T))$ and $\mathbf{Z}_t(t_0) = (\mathbf{X}_t^T, \dfrac{t-t_0}{hT}\mathbf{X}_t^T)^T$.

Define:

$\boldsymbol{\alpha}_{Tj}(t_0) = h_T^2\boldsymbol{\Gamma}^{-1}\sum_{t=1}^T \varepsilon_{tj}^*\mathbf{X}_t K(\dfrac{t-t_0}{hT})$,

$\mathbf{R}_{Tj}(t_0) = h_T^2\sum_{t=1}^T \boldsymbol{\Gamma}^{-1}[\boldsymbol{\Phi}_{0j}^*\mathbf{X}_t - \boldsymbol{\beta}_j(t_0)^T\mathbf{Z}_t(t_0)]\mathbf{X}_t K(\dfrac{t-t_0}{hT})$,

$R_{T1}^j = \sum_{t=1}^T \varepsilon_{tj}^*\mathbf{R}_{Tj}(t)^T\mathbf{X}_t$,

$R_{T2}^j = \sum_{t=1}^T \boldsymbol{\alpha}_{Tj}(t)^T\mathbf{X}_t\mathbf{X}_t^T\mathbf{R}_{Tj}(t)$,

$R_{T3}^j = \dfrac{1}{2}\sum_{t=1}^T \mathbf{R}_{Tj}^T(t)\mathbf{X}_t\mathbf{X}_t^T\mathbf{R}_{Tj}(t)$.

Definition 3.1: (Definition 1 in de Jong(1987)). For each $j = 1, ...k$, W_T^j is called clean if the conditional expectations of W_{stj} vanish:

$$E(W_{stj}|\mathbf{X}_s) = 0, a.s.$$

for all $s, t \leq T$.

Lemma 3.1: (Lemma 7.2 in Fan, Zhang and Zhang(2001)). Under Assumption (B), for each $j = 1, 2, ...k$, as $h \longrightarrow 0$, $Th \longrightarrow \infty$. We have:

$$R^j_{T1} = \sqrt{T}h^2 R^j_{T10} + O(hT^{-1/2}),$$

$$R^j_{T2} = \sqrt{T}h^2 R^j_{T20} + O(hT^{-1/2}),$$

$$R^j_{T3} = Th^4 R^j_{T30} + O(h^3).$$

Furthermore, for any $\delta > 0$, for $j = 1, 2...k$, there exists $M_j > 0$, s.t:
$P(|\frac{R^j_{Ti}}{\sqrt{T}h^2}| > M_j) \leq \delta$, for $i = 1, 2$ and $P(|\frac{R^j_{T3}}{Th^4}| > M_j) \leq \delta$.

Using Lemma 3.1, we can easily show the following lemma.

Lemma 3.2: (Lemma 7.3 in Fan, Zhang and Zhang(2001)). Let $\hat{\Phi}(t/T)$ be the local linear estimator $\hat{P}$ we derived from Lemma 3.1. Let $\hat{\Phi}^*(t/T) \equiv Q\hat{\Phi}(t/T) = (\hat{\Phi}^{T*}_1(t/T), ..., \hat{\Phi}^{T*}_k(t/T))^T$, then under the assumption (B), uniformly for $t_0 \in (0, T)$, for each $j = 1, 2...k$, we have:

$$\hat{\Phi}^*_j(t_0/T) - \Phi_j(t_0/T) = (\alpha_{Tj}(t_0) + R_{Tj}(t_0))(1 + o_p(1)),$$

where $\alpha_{Tj}(t_0)$ and $R_{Tj}(t_0)$ are defined earlier. Again, we define:
$U_{Tj} = h^2_T \sum^T_{t,s=1} \varepsilon^*_{tj}\varepsilon^*_{sj}\mathbf{X}^T_s\Gamma^{-1}\mathbf{X}_t K(\frac{t-s}{hT})$,
$V_{Tj} = h^4_T \sum^T_{t,s=1} \varepsilon^*_{tj}\varepsilon^*_{sj}\mathbf{X}^T_t[\sum^T_{l=1}\Gamma^{-1}\mathbf{X}_l\mathbf{X}^T_l\Gamma^{-1}K(\frac{t-l}{Th})K(\frac{s-l}{Th})]\mathbf{X}_s$.

Lemma 3.3: (Lemma 7.4 in Fan ,Zhang and Zhang(2001)). Under Assumption (B), assume $\varepsilon_t \sim \mathcal{N}(0, \Sigma_0)$, where $Q\Sigma_0 Q^T \equiv (\sigma^2_{ij})^k_{i,j=1}$. As $h \longrightarrow 0$, $Th^{3/2} \longrightarrow \infty$, for $j = 1...k$, we have:

$U_{Tj} = \frac{D}{h}K(0)\sigma_{jj}^2 + \frac{1}{T}\sum_{s\neq t}^{T} \varepsilon_{sj}^* \varepsilon_{tj}^* \mathbf{X}_t^T \Gamma^{-1} \mathbf{X}_s K_h(s-t) + o_p(h^{-1/2})$,
$V_{Tj} = \frac{D}{h}\nu_0 \sigma_{jj}^2 + \frac{2}{Th}\sum_{s<t}^{T} \varepsilon_{sj}^* \varepsilon_{tj}^* \mathbf{X}_s^T \Gamma^{-1} K * K(\frac{s-t}{hT})\mathbf{X}_t + o_p(h^{-1/2})$
with $K_h(\cdot) = \frac{1}{h}K\left(\frac{\cdot}{hT}\right)$.

Proof of Theorem 3.1: Firstly, we show that:

$$\begin{aligned}
\frac{R\widetilde{S}S_a}{T} &= \frac{1}{T}\sum_{t=1}^{T} e_{t1}^T \Sigma^{-1} e_{t1} \\
&= trace(\frac{\sum_{t=1}^{T} e_{t1}^T \Sigma^{-1} e_{t1}}{T}) \\
&= \frac{1}{T}\sum_{t=1}^{T} trace(e_{t1}^T e_{t1}^T \Sigma^{-1}) \\
&= \frac{1}{T} trace([\sum_{t=1}^{T} e_{t1} e_{t1}^T]\Sigma^{-1}) \\
&= \frac{1}{T} trace([T-1]\hat{\Sigma}\Sigma^{-1}) + o_p(1) \\
&= \frac{T-1}{T} trace(I_k) + o_p(1) \\
&= k + o_p(1).
\end{aligned}$$

Knowing that $\hat{\Sigma} = \frac{1}{T-1}\sum_{t=1}^{T} e_{t1} e_{t1}^T$.
Secondly, by the definition, we obtain: for each $j = 1...k$,

$$\begin{aligned}
-\lambda_{Tj}(\Phi_0)\sigma_{jj}^2 = &-h_T^2 \sum_{l=1}^{T} \varepsilon_{lj}^* [\sum_{t=1}^{T} \varepsilon_{tj}^* \mathbf{X}_t^T \Gamma^{-1} \mathbf{X}_l K(\frac{t-t_0}{hT})] \\
&+ \frac{1}{2}h_T^4 \sum_{l=1}^{T}\sum_{t=1}^{T}\sum_{s=1}^{T} \varepsilon_{tj}^* \varepsilon_{sj}^* \mathbf{X}_t^T \Gamma^{-1} \mathbf{X}_l \mathbf{X}_l^T \mathbf{X}_s \Gamma^{-1} K(\frac{t-l}{hT})K(\frac{s-l}{hT}) \\
&- R_{T1}^j + R_{T2}^j + R_{T3}^j + O_p(\frac{1}{Th^2}).
\end{aligned}$$

Thus, we apply Lemma 3.1, Lemma 3.2 and Lemma 3.3. We find out:

$$-\lambda_{Tj}(\Phi_0) = \mu_T + d_{1Tj} - \frac{1}{2\sqrt{h}} W_T^j + o_p(h^{-1/2}).$$

where $W_T^j = \dfrac{\sqrt{h}}{T\sigma_{jj}^2} \sum_{s\neq t}^T \varepsilon_{tj}^* \varepsilon_{sj}^* [2K_h(s-t) - K_h * K_h(s-t)] \mathbf{X}_s^T \Gamma^{-1} \mathbf{X}_t$.

Now, we need to show that for all $j = 1, ...k$,

$$W_T^j \xrightarrow{L} N(0, w)$$

where $w = 2D\|2K - K * K\|_2^2$. Define:

$$W_{stj} = \frac{\sqrt{h}}{T} c_T(s,t) \varepsilon_{sj}^* \varepsilon_{tj}^* / \sigma_{jj}^2,$$

for $1 \leq s < t \leq T$ with $c_T(s,t)$ can be written as:

$$c_T(s,t) = b_1(s,t) + b_2(s,t) - b_3(s,t) - b_4(s,t)$$

where $b_1(s,t) = 2K_h(s-t)\mathbf{X}_s^T \Gamma^{-1} \mathbf{X}_t$, $b_2(s,t) = b_1(t,s)$;

$b_3(s,t) = K_h * K_h(s-t)\mathbf{X}_s^T \Gamma^{-1} \mathbf{X}_t$, $b_4(s,t) = b_3(t,s)$.

Hence $W_T^j = \sum_{s<t}^T W_{stj}$, for $j = 1...k$.

In order to employ Proposition 3.2 in de Jong(1987), we need to check the following conditions:

(1) W_T^j is clean.

(2) $var(W_T^j) \longrightarrow w$. as $T \longrightarrow \infty$.

(3) G_I^j is of smaller order than $var(W_T^j)$.

(4) G_{II}^j is of smaller order than $var(W_T^j)$.

(5) G_{IV}^j is of smaller order than $var(W_T^j)$.

where

$G_I^j = \sum_{1 \le s < t \le T} E(W_{stj}^4)$,

$G_{II}^j = \sum_{1 \le s < t < l \le T} [E(W_{stj}^2 W_{slj}^2) + E(W_{tsj}^2 W_{tlj}^2) + E(W_{lsj}^2 W_{ltj}^2)]$,

$G_{IV}^j = \sum_{1 \le s < t < l < u \le T} [E(W_{stj} W_{slj} W_{utj} W_{ulj}) + E(W_{stj} W_{suj} W_{ltj} W_{luj}) + E(W_{slj} W_{suj} W_{tlj} W_{tuj})]$.

Now we check each of the conditions above.

Condition (1) follows straightforwardly from the definition.

To verify (2), we notice that: $var(W_T^j) = \sum_{s<t}^T E(W_{stj}^2)$. Denote:

$K(v, m) = K * \cdots * K(v)$ as the m-th convolution of $K(\cdot)$ at v for $m = 1, 2....$ Therefore it follows that:

$$E[c_T^2(s,t)\varepsilon_{sj}^{*2}\varepsilon_{tj}^{*2}] = \frac{D\sigma_{jj}^4}{h}[16K(0,2) - 16K(0,3) + 4K(0,4)](1 + O(h))$$

which leads to: $w = 2D \int [2K(x) - K * K(x)]^2 dx = 2D\|2K - K * K\|_2^2$.

Condition (3) is satisfied by noting that for each $j = 1, 2...k$,

$$E[b_1(1,2)\varepsilon_{1j}^*\varepsilon_{2j}^*]^4 = O(h^{-3})b_3(1,2)\varepsilon_{1j}^*\varepsilon_{2j}^*]^4 = O(h^{-2}).$$

which implies that $E(W_{12j}^4) = \dfrac{h^2}{T^4} O(h^{-3})$. Thus, $G_I^j = O(T^{-2}h^{-1}) = o(1)$.

Condition (4) is verified by the following calculation:

$$E(W_{12j}^2 W_{13j}^2) = O(E(W_{12j}^4)) = O(T^{-4}h^{-1})$$

which gives that: $G_{II}^j = O(T^{-1}h^{-1}) = o(1)$ for all $j = 1, 2...k$.

To prove condition (5), it suffices to compute the term $E(W_{12j}W_{23j}W_{34j}W_{41j})$. By direct calculations:

$E[b_1(1,2)b_1(2,3)b_1(3,4)b_1(4,1)\varepsilon_{1j}^{*2}\varepsilon_{2j}^{*2}\varepsilon_{3j}^{*2}\varepsilon_{4j}^{*2}] = O(h^{-1})$

$E[b_1(1,2)b_1(2,3)b_1(3,4)b_3(4,1)\varepsilon_{1j}^{*2}\varepsilon_{2j}^{*2}\varepsilon_{3j}^{*2}\varepsilon_{4j}^{*2}] = O(h^{-1})$

$E[b_1(1,2)b_1(2,3)b_3(3,4)b_3(4,1)\varepsilon_{1j}^{*2}\varepsilon_{2j}^{*2}\varepsilon_{3j}^{*2}\varepsilon_{4j}^{*2}] = O(h^{-1})$

$E[b_1(1,2)b_3(2,3)b_3(3,4)b_3(4,1)\varepsilon_{1j}^{*2}\varepsilon_{2j}^{*2}\varepsilon_{3j}^{*2}\varepsilon_{4j}^{*2}] = O(h^{-1})$

$E[b_3(1,2)b_3(2,3)b_3(3,4)b_3(4,1)\varepsilon_{1j}^{*2}\varepsilon_{2j}^{*2}\varepsilon_{3j}^{*2}\varepsilon_{4j}^{*2}] = O(h^{-1})$

and similarly for the other terms. Hence,

$$E(W_{12j}W_{23j}W_{34j}W_{41j}) = T^{-4}h^2O(h^{-1}) = O(T^{-4}h)$$

which leads to: $G_{IV}^j = O(T^4T^{-4}h) = O(h) = o(1)$.

By now, we have shown for each $j = 1, 2...k$, we have, under H_0:

$$\sigma_T^{-1}(\lambda_{Tj}(\Phi_0) - \mu_T + d_{1Tj}) \xrightarrow{L} \mathcal{N}(0,1)$$

where $\sigma_T^2 = \frac{D}{2h}\int (2K(x) - K * K(x))^2 dx$,
$\mu_T = \frac{D}{2h}(2K(0) - \inf K^2(x)dx)$,
$d_{1Tj} = \sigma^{-2}[Th^4R_{T30}^j - T^{1/2}h^2(R_{T10}^j - R_{T20}^j)] = O_p(Th^4 + \sqrt{T}h^2)$.

From the definition, we get our GQLR test statistic, under H_0:

$$\begin{aligned}
\lambda_T(\Phi_0) &\approx \frac{T}{2}\frac{R\tilde{S}S_0 - R\tilde{S}S_a}{R\tilde{S}S_a} \\
&\approx \frac{R\tilde{S}S_0 - R\tilde{S}S_a}{2k} \\
&= \frac{1}{2k}\sum_{j=1}^{k}\lambda_j(RSS_0 - RSS_a)(j) \\
&= \frac{1}{k}\sum_{j=1}^{k}\lambda_j\sigma_{jj}^2\lambda_{Tj}(\Phi_0)
\end{aligned}$$

Since $\lambda_{Tj}(\Phi_0) - \mu_T + d_{1Tj} \xrightarrow{L} \mathcal{N}(0, \sigma_T^2)$. Thus, for each $j = 1, ...k$:

$$E(\lambda_j\sigma_{jj}^2\lambda_{Tj}(\Phi_0) - \lambda_j\sigma_{jj}^2\mu_T + \lambda_j\sigma_{jj}^2 d_{1Tj}) = 0$$

Let

$$\mu_T^* = \frac{\mu_T}{k} \sum_{j=1}^{k} \lambda_j \sigma_{jj}^2,$$
$$d_{1T}^* = \frac{1}{k} \sum_{j=1}^{k} \lambda_j \sigma_{jj}^2 d_{1Tj}$$

then $E(\lambda_T(\Phi_0) - \mu_T^* + d_{1T}^*) = 0$.

We already shown that $var(\lambda_{Tj}(\Phi_0)) = \sigma_T^2(1 + O(1))$.

Now, we focus on the variance of $\lambda_T(\Phi_0)$, we have:

$$\begin{aligned} var[\lambda_T(\Phi_0)] &= var[\frac{1}{k} \sum_{j=1}^{k} \lambda_j \sigma_{jj}^2 \lambda_{Tj}(\Phi_0)] \\ &= \frac{\sigma_T^2}{k^2} \sum_{j=1}^{k} \lambda_j^2 \sigma_{jj}^4 + \frac{2}{k^2} \sum_{1 \le i < j \le k} cov(\lambda_j \sigma_{jj}^2 \lambda_{Tj}(\Phi_0), \lambda_i \sigma_{ii}^2 \lambda_{Ti}(\Phi_0)) \end{aligned}$$

Since $-\lambda_{Tj}(\Phi_0) = \mu_T + d_{1Tj} - \frac{1}{2\sqrt{h}} W_T^j + o_p(h^{-1/2})$, for $i < j$, we obtain:

$$\begin{aligned} cov(\lambda_i \sigma_{ii}^2 \lambda_{Ti}(\Phi_0), \lambda_j \sigma_{jj}^2 \lambda_{Tj}(\Phi_0)) &= \frac{\lambda_i \lambda_j \sigma_{ii}^2 \lambda_{jj}^2}{4h} cov(W_T^i, W_T^j) \\ &= \frac{\lambda_{ij} \sigma_{ii}^2 \lambda_{jj}^2}{4h} E(W_T^i W_T^j), \end{aligned}$$

Similar with the calculation of $var(W_T^j)$, we obtain:

$$\begin{aligned} E(W_T^i W_T^j) &= E[(\sum_{s<t} W_{ti})(\sum_{s<t} W_{stj})] \\ &= \sum_{1 \le s < t \le T} E(W_{sti} W_{stj}) \end{aligned}$$

$$= \sum_{1 \le s < t \le T} E[\frac{hc_T^2(s,t)}{T^2 \sigma_{ii}^2 \sigma_{jj}^2} \varepsilon_{si}^* \varepsilon_{ti}^* \varepsilon_{sj}^* \varepsilon_{tj}^*]$$

We note that:

$$E(c_T^2(s,t) \varepsilon_{si}^* \varepsilon_{ti}^* \varepsilon_{sj}^* \varepsilon_{tj}^*) = \frac{D\sigma_{ij}^4}{h}[16K(0,2) - 16K(0,3) + 4K(0,4)](1 + O(h)).$$

Therefore, we have:

$$var(\lambda_T(\Phi_0) = \frac{\sigma_T^2}{k^2}[\sum_{j=1}^{k} {}_j^2 \sigma_{jj}^4 + 2\sum_{i<j}^{k} \lambda_i \lambda_j \sigma_{ij}^4](1 + O(1)).$$

Denote: $\sigma^{*2} \equiv \frac{\sigma_T^2}{k^2}[\sum_{j=1}^k \lambda_j^2 \sigma_{jj}^4 + 2\sum_{i<j}^k \lambda_{ij} \sigma_{ij}^4]$. Then:

$$var(\lambda_T(\Phi_0)) \longrightarrow \sigma^{*2}.$$

where $\sigma_T^2 = \frac{2D}{h} \int (K(x) - \frac{1}{2} K * K(x))^2 dx$.

Notice that, $\lambda_T(\Phi_0) - \mu_T^* + d_{1T}^*$ is clean. In order to apply Proposition 3.2 in de Jong (1987) again, it remains to check:

(1) F_I^{ij} is of smaller order than $E(W_T^i W_T^j)$

(2) F_{II}^{ij} is of smaller order than $E(W_T^i W_T^j)$

(3) F_{IV}^{ij} is of smaller order than $E(W_T^i W_T^j)$

where $F_I^{ij} = \sum_{1 \le s < t \le T} E(W_{sti}^2 W_{stj}^2)$,

$F_{II}^{ij} = \sum_{1 \le s < t < l \le T} [E(W_{sti} W_{sli} W_{stj} W_{slj}) + E(W_{tsi} W_{tli} W_{tsj} W_{tlj}) + E(W_{lsi} W_{lti} W_{lsj} W_{ltj})]$,

$F_{IV}^{ij} = \sum_{1 \le s < t < l < u \le T} [E(W_{sti} W_{slj} W_{utj} W_{uli}) + E(W_{sti} W_{suj} W_{ltj} W_{lui}) + E(W_{sli} W_{suj} W_{ltj} W_{lui}) + E(W_{sli} W_{suj} W_{tlj} W_{tui})]$.

Condition (1) holds because $E(W_{12i}^2 W_{12j}^2) = O(E(W_{12i}^4)) = O(T^{-4}h^{-1})$, Hence, $F_I^{ij} = O(T^{-2}h^{-1}) = o(1)$.

To prove (2), note that:

$$E(W_{12i}W_{13i}W_{12j}W_{13j}) = O(E(W_{12i}^2 W_{13i}^2))$$
$$= O(E(W_{12i}^4)) = O(T^{-4}h^{-1})$$

Thus, $F_{II}^{ij} = O(T^{-1}h^{-1}) = O(1)$.

To prove (3), it suffices to calculate the term $E(W_{12i}W_{23j}W_{34i}W_{41j})$. By straightforward calculations:

$E[b_1(1,2)b_1(2,3)b_1(3,4)b_1(4,1)\varepsilon_{1i}^*\varepsilon_{2i}^*\varepsilon_{3i}^*\varepsilon_{4i}^*\varepsilon_{1j}^*\varepsilon_{2j}^*\varepsilon_{3j}^*\varepsilon_{4j}^*] = O(h^{-1})$,

$E[b_1(1,2)b_1(2,3)b_1(3,4)b_3(4,1)\varepsilon_{1i}^*\varepsilon_{2i}^*\varepsilon_{3i}^*\varepsilon_{4i}^*\varepsilon_{1j}^*\varepsilon_{2j}^*\varepsilon_{3j}^*\varepsilon_{4j}^*] = O(h^{-1})$,

$E[b_1(1,2)b_1(2,3)b_3(3,4)b_3(4,1)\varepsilon_{1i}^*\varepsilon_{2i}^*\varepsilon_{3i}^*\varepsilon_{4i}^*\varepsilon_{1j}^*\varepsilon_{2j}^*\varepsilon_{3j}^*\varepsilon_{4j}^*] = O(h^{-1})$,

$E[b_1(1,2)b_3(2,3)b_3(3,4)b_3(4,1)\varepsilon_{1i}^*\varepsilon_{2i}^*\varepsilon_{3i}^*\varepsilon_{4i}^*\varepsilon_{1j}^*\varepsilon_{2j}^*\varepsilon_{3j}^*\varepsilon_{4j}^*] = O(h^{-1})$,

$E[b_3(1,2)b_3(2,3)b_3(3,4)b_3(4,1)\varepsilon_{1i}^*\varepsilon_{2i}^*\varepsilon_{3i}^*\varepsilon_{4i}^*\varepsilon_{1j}^*\varepsilon_{2j}^*\varepsilon_{3j}^*\varepsilon_{4j}^*] = O(h^{-1})$

Then, $E(W_{12i}W_{23j}W_{34i}W_{41j}) = \dfrac{h^2}{T^4}O(h^{-1}) = O(T^{-4}h)$. yielding

$$F_{IV}^{ij} = O(h) = o(1),$$

Therefore, we have shown that $var(\lambda_T(\Phi_0))$ has been dominated by σ^{*2}. Hence,

$$\sigma^{*-1}(\lambda_T(\Phi_0) - \mu_T^* + d_{1T}^*) \xrightarrow{D} \mathcal{N}(0,1)$$

This finishes the proof. □.

Proof of Theorem 3.2: Theorem 3.2 is one special case of Theorem 3.1 when Φ_0^* under H_0 is a vector of constants. So, with the same notation as in the proof of Theorem 3.1, we have for each $j = 1, ...k$, $\Phi_0^{*"} = 0$. Hence, $R_{T10}^j = R_{T20}^j = R_{T30}^j = 0$ which leads to each $d_{1Tj} = 0$ and $d_{1T}^* = 0$. The rest of the proof is the same as the proof Theorem 3.1. □.

Proof of Theorem 3.3: Under H_a and Assumption (B), applying Theorem 3.1, we have: for each $j = 1, ...k$,

$$-\lambda_{Tj}(\Phi_0) = -\mu_T + \nu_{Tj} - d_{2T} - [\frac{1}{2\sqrt{h}} W_T^j + \sum_{t=1}^T G^T(u)(I_k \otimes \mathbf{X}_t)\varepsilon_t/(\sigma_{jj}^2)] + o_p(h^{-1/2}).$$

where

$$\nu_{Tj} = \frac{Th^4}{8\sigma_{jj}^2} E\{\mathbf{\Delta}^{"T}(u)[(\mathbf{X}_t\mathbf{X}_t^T) \otimes I_k]\mathbf{\Delta}^{"}(u)\}\omega_0 \text{ and}$$

$$d_{2T} = \frac{T}{2} E\{\mathbf{G}^T(u)[(\mathbf{X}_t\mathbf{X}_t^T) \otimes I_k]\mathbf{G}(u)\}.$$

with μ_T and W_T^j are defined in the proof of Theorem 3.1. The rest of the proof is similar to the proof of Theorem 3.1. The details are omitted. $\square$.

CHAPTER 4: CONCLUSION

In this dissertation, I propose some new test procedures, called generalized quasi-likelihood ratio test, to test some hypotheses for multivariate time-varying coefficient regression model for time series data, such as testing whether coefficients are indeed time-varying or of some specific functional form.

First of all, I use local linear technique to estimate the nonparametric coefficients functions and derive the explicit representation of the estimators. I also develop the theoretical asymptotic distribution of the estimators I derive.

Secondly, I propose the new test statistic which is built based on the comparison of the quasi-likelihood under between null and alternative hypotheses. I give the theoretical asymptotic null and alternative distributions. The Monte Carlo simulations are conducted to illustrate the power of the proposed test procedure and an application to a real data set is presented as well.

There are still many interesting research topics related to this dissertation which deserve further investigation. First, one may release the stationary or mixing conditions. I only focus on the asymptotic result under stationary time series data setting. Secondly, the generalized quasi-likelihood ratio test statistic can be extended to other models. For example, additive models, predictive regression models and so on. Last but not the least,few paper available in literature about multivariate time-varying coefficient models under nonstationary time series setting due to the difficulty of deriving explicit representation for the nonstationary data. All of the above issues can should be given a lot of attention as a future research topic.

REFERENCES

[1] Barnes, M and Hughes, A.2002. A Quantile Regression Analysis of the Cross Section of Stock Market Returns, working paper, Federal Reserve Bank of Boston.

[2] Bassett, G and Chen, H.2001. Portfolio Style: Return-Based Attribution Using Quantile Regression, Empirical Economics, Springer-Verlag, pp. 1405C1441.

[3] Cai, Z. (2002a). Regression quantile for time series. Econometric Theory 18, 169-192.

[4] Cai, Z., Fan, J., Yao, Q., 2000. Functional-coefficient regression models for nonlinear time series. Journal of the American Statistical Association 95, 941-956.

[5] Cai, Z., Xu, X., 2008. Nonparametric quantile estimations for Dynamic smooth coefficient models. Journal of the American Statistical Association. 103, 1595-1608.

[6] Cai, Z., Das, M., Xiong, H.,Wu, X.,2006. Functional coefficient instrumental variables models. Journal of Econometrics. 133, 207-241.

[7] Cai, Z., Xiao, Z., Semiparametric Quantile Regression Estimation in Dynamic Models with partially Varying Coefficients. Journal of Econometrics. To be coming.

[8] Chaudhuri, P., 1991. Nonparametric quantile estimates of regression quantiles and their local Bahadur representation. The Annals of Statistics 19, 760-777.

[9] Cleveland, W. S., Grosse, E. and Shyu, W. M. (1991). Local regression models. In Statistical Models in S (Chambers, J. M. and Hastie, T. J., eds), 309C376. Wadsworth & Brooks, Pacific Grove.

[10] Engle, R. andManganelli, S. 2004. CAViaR: Conditional Autoregressive Value at Risk by Regression Quantiles, Journal of Business and Economic Statistics.22(4), 367-381

[11] Fan, J., Yao,Q., and Tong. H.(1996),"Estimation of Conditional Densities and Sensitivity Measures in Nonlinear Dynamical Systems," Biometrika,83,189-206.

[12] Fan, J., Zhang, C., Zhang, J., 2001. Generalized Likelihood Ratio Statistics and Wilks Phenomenon. The Annuals of Statistics.29,153-193.

[13] Hastie, T. J. and Tibshirani, R. J. (1993). Varying-coefficient models. J. Roy. Statist. Soc. B. 55 757C796.

[14] He, X. Shi,P., 1996. Bivariate tensor-product B-splines in a partly linear model. Journal of Multivariate Analysis 58, 162-181.

[15] He,X., Ng, P., and Portnoy, S. 1998. Bivariate Quantile Smoothing Splines. Journal of the Royal Statistical Society. Ser. B, 60. 537-550.

[16] Hencricks, W. and Koenker, R. (1992). Hierarchical spline models for conditional quantiles and the demand for electricity. J. Amer. Statist. Assoc. 87 58C68.

[17] Honda,T. 2004. Quantlie Regression in Varying Coedfficient Models. Journal of Statistical Planning and Inference, 121, 113-125.

[18] Cai, Z., Fan, J. and Li, R. (2000). Efficient estimation and inferences for varying-coefficient models. J. Amer. Statist. Assoc. 95 888C902. MR1804446

[19] Horowitz, J. L.,and Lee, S. 2005. Nonparametric Estimation of an Additive Quantile Regression Model.Journal of the American Statistical Association, 100, 1238-1249.

[20] Koenker, R. and G. Bassett 1982a Robust Tests for Heteroscedasticity Baxed on Regression Quantiles, Econometrica, 50, 43-61.

[21] Koenker, R. and G. Bassett. 1982b. Test of Linear Hypotheses and L_1 Estimation, Econometrica, 50, 1577-1584.

[22] Koenker, R., 2004. Quantreg: An R Package for Quantile Regression and Related Methods. http://cran.r-project.org.

[23] Koenker, R., 2005. Quantile Regression. Econometric Society Monograph Series, Cambridge University Press, New York.

[24] Koenker, R., Bassett, G.W.,1978. Regression Quantiles. Econometrica 46, 33-50.

[25] Koenker, R., and Xiao, Z.(2004),"Unit Root Quantile Autoregression Inference," Journal of the American Statistical Association,99,775-787.

[26] Komunjer, I., 2005. Quasi-maximum likelihood estimation for conditional quantiles. Journal of Econometrics 128, 137-164.

[27] Wu, T., Yu, K. and Yu,Y. 2010. Single-index quantile regression. Journal of Multivariate Analysis. 101, 1607-1621.

[28] Zhang, C. (2000). Topics in Generalized Likelihood Ratio Test. Phd dissertation, The University of North Carolina at Chapel Hill.

[29] Rose, S.J. (1992). Social Stratification in the United States. The New Press, New York.

[30] El Bantli, F. and M. Hallin (1999), L_1 Estimation in linear models with heterogeneous white noise, Statistics and Probability Letters, 45, 305-315.

[31] Horowitz, J. L. and Spokpiny, V. G. (2002). An adaptive, rate-optimal test of linearity for median regression models. J. Amer. Statist. Assoc. 97 822C835. MR1941412

[32] He, X. and Zhu, L.X. (2003). A lack-of-fit test for quantile regression. J. Amer. Statist. Assoc. 98 1013C1022. MR2041489

[33] Gourieroux, C., Monfort, A. and Renault, E. (1987), Consistent M-Estimators in a Semi- Parametric Model, Document de Travail de lINSEE, 8706.

[34] Koenker, R. and J. Machado. 1999, Goodness of Fit and Related Inference Processes for quantile Regression. Journal of the Royal Statistical Society, Series B,66, 145-163

[35] Roncheti, E. 1985, Robust Model Selection in Regression. Statistics &Probability Letters, 3, 21-23.

[36] Harrison and Rubinfeld (1978), "Hedonic prices and the demand for clean air",Journal of Environmental Economics and Management, Vol. 5, pp. 81-102.

[37] Fan, J. and Gijbels, I.(1996). Local Polynomial Modeling and Its Applications. Chapman and Hall, London.

[38] Fan,J. and Jiang,J (2005). Nonparametric inferences for additive models. Journal of the American Statistical Association, 100,891-907

[39] Tsay, R. S. (1989).Testing and modeling threshold autoregressive processes. Journal of the American Statistical Association 84, 231-240.